World Productivity Growth: A Model Averaging Approach[*]

Meryem Duygun
Nottingham University Business School
Nottingham, UK
Meryem.Duygun@nottingham.ac.uk

Anders Isaksson
United Nations Industrial Development Organization
Vienna, Austria
A.Isaksson@unido.org

Jiaqi Hao
Credit Scorecard and Portfolio Management/ATB Financial
Edmonton, Alberta
jhao@atb.com

Robin C. Sickles
Rice University/Economics
Houston, Texas
rsickles@rice.edu

This version June 8, 2017

Abstract

Policy makers and the economic researchers who provide them estimates of economic activity need to have an informative and scientifically-based method to develop a consensus estimate for the most basic of the productivity measures, total factor productivity (TFP) growth. We discuss methods to combine the various estimates based on different empirical specifications that model and estimate productivity growth. We also discuss the various econometric approaches used in the profession to estimate productivity growth. Our focus is on world TFP growth.

Keywords: Productivity; Panel Data; Economic Growth; Time Varying Unobservable Effects

JEL Classification Numbers: D24, C23, O47.

1 Introduction

The measure of nations' productivity is of great importance for both academics and policymakers in assessing performance and planning future economic roadmaps. Strategic decisions require robust indicators that receive broad consensus from various parties involved in the decision-making process. The importance of such consensus is further amplified when monitoring bodies, and lenders such as international organizations or financial and development institutions (e.g., IMF, World Bank, UN and Regional Development Banks), are involved. This underscores the long debated and largely unanswered question in the literature: on which approach do we rely?

A common measure of productivity is the Total Factor Productivity (TFP) Index. However, the existence of numerous methods and models involved in TFP measurement make it a cumbersome and puzzling process for policy makers. This paper proposes a practical yet robust approach that provides an appealing solution to academics and policy makers alike in their pursuit for a "consensus" productivity indicator. We propose that instead of relying on a single approach, one should gather information from a set of measurement methods and construct a single productivity indicator that is backed by robust methodological techniques.

To demonstrate our approach, we apply it to a unique data set that covers a large set of countries across the globe, namely the World Productivity Database (WPD) developed by the United Nations Industrial Development Organization (UNIDO). One of the purposes of WPD is to speak to the many approaches to TFP measurement and provide productivity analysts across the globe with TFP estimates based on numerous methods, production function specifications, functional forms, different capital stock and labor input measures, and much more. The arsenal of approaches available to the researcher today, many of which are reflected in WPD, is manifold and technically advanced and invites researchers to provide comparisons of results obtained from applying several methods. Unfortunately, this approach is seldom the case, as analysts tend to resort to one method only.

The advantage of having a rich toolkit is, of course, a potential increase in accuracy with which we are able to measure productivity performance. However, it also constitutes an acknowledgement of uncertainties involved in modeling the productivity measure. A notable disadvantage is that these different measurement methods yield a range of estimates with sometimes very wide dispersion or, in the worst case, conflicting results. Furthermore, all models may be subject to misspecification of unknown form, e.g., researchers might have different information sets. Moreover, models may be affected differently by structural breaks caused by institutional change or technological development, to name but a few possible reasons leading to variation in TFP measurement.

However commendable the work of WPD may be, it is still silent on the issue of what the "correct" productivity estimate is. Importantly, the question is how policy makers are going to be able to make the "right" choice from available alternative approaches. The approach presented in this paper is a step towards resolving this conundrum.

To demonstrate the strength of our approach, the analysis presented, *inter alia*, includes results at the aggregate world level, compares the performance of six country groups at different stages of development, and decomposes TFP growth into change in technical efficiency and innovation. Such decomposition provides policy makers with a richer and more detailed basis for policy making. Importantly, the analysis introduces a comparison of our consensus estimates with those provided by common approaches such as growth accounting, pooled and panel regression analysis, and data envelopment analysis. Our consensus estimates fare well in comparison and we conclude that it may be advisable to combine estimates in order to make the best conclusion based on all the available information.

The paper provides a brief discussion of panel data and productivity analysis in applied economic modeling. We discuss a variety of modeling scenarios and justifications for them based on classical economic theory and on more recent advances in production modeling that formulate methods to decompose productivity growth based on a Solow-type residual (Solow, 1957) into innovation and catch-up, the latter referred to as technical efficiency change in the stochastic and non-parametric frontier literature. We point to a number of innovations contributed to the panel data literature by those working in the stochastic frontier productivity discipline. In that literature the focus has been on the interpretation of relative temporal heterogeneity between production units (firms, countries,

etc.) as a measure of relative technical efficiency in the use of the frontier technology. Our paper has an aggregate productivity perspective, focusing on country level productivity, as it better motivates and displays the strong intellectual parallels between the efficiency literature, the economic growth and development literature, and the literature on panel data econometrics.

The paper is organized in the following way. We first discuss how productivity growth typically has been measured in classical productivity studies. We then briefly discuss how innovation and catch-up can be distinguished empirically. We discuss the econometric methods that accomplish this. In the Appendix we also provide a set of Monte Carlo simulations that assess the performance of the model averaging estimators we employ in our empirical work. We detail the motivation for and the methods used in developing a consensus estimate from the competing model estimates. We apply these methods to develop consensus estimates for world TFP growth from 1970 to 2000 and then conclude.

2 Measuring Productivity and its Growth

2.1 Solow residual-based methods

In order to account for changing input (X_i) mix, modern index number analyses utilize a measure of total factor productivity (TFP) for a single output (Y) technology that in its simplest form is a ratio of output to a weighted sum of inputs. A useful construct for a single output technology is thus :

$$TFP = \frac{Y}{\sum a_i X_i}.$$ (1)

Detailed discussion of the properties and formulation of productivity measures for single and multiple output technologies are plentiful in the productivity literature. One relatively early and informative discussion was undertaken by productivity pioneers Jorgenson and Griliches (1972). Many others have preceded and followed their work. The literature is deep and extensive and we do not attempt to list the relevant references in this paper.

Solow's residual-based measure is based on the Cobb-Douglas production function with constant returns to scale, $Y = AX_L^\alpha X_K^{1-\alpha}$ and given by:

$$TFP = \frac{Y}{X_L^\alpha X_K^{1-\alpha}}.$$ (2)

Cost-minimization allows one to describe the TFP growth index in terms of expenditure shares:

$$T\dot{F}P = \frac{dY}{Y} - \left[\alpha \frac{dX_L}{X_L} + (1-\alpha)\frac{dX_K}{X_K}\right]$$

where the non-negative parameter α is the input expenditure share for labor. When multiple outputs exist, TFP can also be described as a ratio of an index number describing aggregate output levels(Y_j) divided by an index number describing aggregate input levels(X_i). As is well-known, since TFP is a function of index numbers, it derives its properties from the aggregator functions on which it is based (see, e.g., Good, Nadiri, and Sickles, 1997).

3 Sources of Economic Growth, the Neoclassical and New Growth Theory Models

Debates among researchers on the primary sources of economic growth and development centered on two basic explanations that are rooted in the decomposition of economic growth sources: factor-accumulation and productivity-growth components. According to Kim and Lau (1994), Young (1992, 1995) and Krugman (1994), rapid economic growth in such emerging areas as East Asia was largely explained by the mobilization of resources. Alternative explanations to the neoclassical

growth model explain economic growth not only in terms of intensive and extensive utilization of input factors but also due to factors that impact the degree to which countries can appropriate the productivity potential of world technical innovations. Again, factors such as governmental industrial policies, trade liberalization policies, and political, religious, and cultural institutions are often viewed as central to the ability of countries to catch-up with a shifting world production possibilities frontier (Grosskopf and Self, 2006)

Stiroh (2001) provides a coherent treatment of neoclassical theory that frames the problem of measuring sources of TFP growth in the context of the neoclassical production $Y = f(K, L, T)$ where variables are indexed by a time subscript. In the typical neoclassical model of production, the production function exhibits constant returns to scale, positive and diminishing returns to each input, with marginal products approaching zero (infinity) as each input goes to infinity (zero). As noted by Stiroh (and many others), long run per capita output growth is exogenously determined by technical change. The neoclassical growth model is not a model that explains long-run growth since technical change, which is the sole determinant of productivity growth, is determined outside the system.

The neoclassical growth model can be modified to address this shortcoming by allowing for technical change to be determined within the model, that is to be endogenously determined. This leads to the "new" growth theory of Romer (1986). In the Romer model, non-diminishing returns to capital were due to external research and development effects. These were treated as spillovers, affecting the stock of knowledge for all firms. In the simple Romer model firms face constant returns to scale to all private inputs. The level of technology A can vary depending on the stock of some privately provided input R (such as knowledge) and the production function is formulated as

$$Y = A(R)f(K, L, R)$$

The sources of the spillover differentiate many contributions to this literature. In the Arrow (1962) model the source of the spillover was learning-by-doing, in the Lucas (1988) model it was the stock of human capital, in the Coe and Helpman (1995) it was trade spillovers. Diao *et al.* (2005) examined the impacts of both a protectionist alternative and shock liberalization and concluded that reduced openness had a negative impact on the overall growth rate due to reduced learning from the foreign spillover. Blazer and Sickles (2009) pursued the spatial effects of learning spillovers in their analysis of the determinants of rapid gains in the productivity of constructing "liberty ships" during World War II.

As pointed out by Abramovitz (1986), Dowrick and Nguyen (1989), and Nelson and Wright (1992), among many others, sources of productivity differences in post WWII industrialized countries can be explained by neoclassical growth models that incorporate knowledge spillovers, technological diffusion, and convergence to a best practice production process (Smolny, 2000). The "new growth theory" implicitly recognizes the role of efficiency in production. There are of course many other explanations for possible technology spillovers. In the productivity and efficiency literature these lead to an asymmetric error term that contains both the classical Solow residual and an additional component reflecting the inability of a firm or country to take advantage of the existing technology. This is simply viewed as a constraint on the unimpeded utilization of the available technology spillovers and thus the asymmetry is due to the firm or country's inability to fully utilize the "frontier" technology (Sickles and Cigerli, 2009). One set of papers that provides an explicit efficiency interpretation for this growth process is Hultberg, Nadiri and Sickles (1999, 2004) and Ahn, Good, and Sickles (2000), which introduce inefficiency into the growth process. Of course, the standard neoclassical model without explicit treatment of efficiency has been used by many authors in examining growth and convergence. An implication of the endogenous growth model is that if a time trend is added to the standard neoclassical production function then the trend must be stochastic. This clearly has implications for stationarity (Reikard, 2005; Bai, Kao, and Ng, 2009). Kneip, Sickles, and Song (2012) have addressed the estimation issues that are associated with specifying endogenous technical change in the presence of arbitrary forms of cross-sectional and time-varyng efficiency change.

3.1 Alternatives to the Neoclassical Growth Model

A more formal rationale for the nonsymmetry in the Solow residual and its interpretation as a measure in the inability to take full advantage of an existing technology is found in the stochastic frontier production literature introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van Den Broeck (1977). The original econometric specifications to address the nonsymmetry of the Solow residual were cross-sectional approaches. These did not lend themselves to a study of country dynamics as it was not possible with these cross-sectional models to identify efficiency differences in different countries and temporal changes in those measures, or catch-up. Panel data extensions to address these shortcomings were introduced by Pitt and Lee (1981) and Schmidt and Sickles (1984), based on random and fixed effect panel models. Time-varying heterogeneity that allowed country effects to vary over time was introduced into these models by Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), and Lee and Schmidt (1993).

The regression-based approaches to estimating sources of time varying and country specific total factor productivity growth utilize panel data methods in specifying time varying technical inefficiency captured by normalized (possibly time-varying) intercepts or fixed effects. Technical inefficiency can also be identified through normalized "within" residuals from error components models with the technical inefficiency effects. Moreover, parametric distributions can be assumed for such panel random effect models and maximum likelihood can be used. For example, a truncated normal distribution with time varying means can be specified as the one-sided error process for technical efficiency (Battese and Coelli, 1992). Cuesta (2000) generalized Battese and Coelli (1992) by allowing each country to have its own time path of technical inefficiency. The assumption of independence between inputs and technical efficiency is problematic as is the incidental parameters problem of MLE when fixed effects are assumed since the number of parameters increases with the sample size.

One can also utilize reduced-form two stage approaches to estimate correlations between inefficiency and measured determinants, such as a country's institutions (Hultberg et al., 1999, 2004). Hultberg *et al.* examined second stage regressions of efficiency on a set of institutional factors and found that variations in these factors explained almost two-thirds of the total variation in country efficiency.

4 Decomposition of Economic Growth-Innovation and Efficiency Change Identified by Regression-based Methods

The focus of this paper is on the econometric, or regression-based, estimation of TFP growth that allows one to separately identify and calculate technological change and efficiency change, and is quite standard. We limit ourselves to standard linear parametric functions in our model averaging analysis as the computational burden is already rather extreme without considering nonparametric and nonlinear specifications of the production function. The general model we consider is:

$$y_{it} = x_{it}\beta + \eta_i(t) + v_{it} \tag{3}$$

where $\eta_i(t)$ represents the country specific fixed effect that may be time varying, x_{it} is a vector of regressors, some of which may be endogenous and correlated with the error v_{it} or the effects $\eta_i(t)$.

Assuming that technical change is appropriable by all countries and that the relative magnitudes of country-specific effects are due to relative inefficiencies between them, TFP growth can be decomposed econometrically into technical change and efficiency change when countries operate at constant-returns-to-scale. When they do not one can also construct a decomposition of TFP growth that also controls for scale effects. We focus on decompositions that obtain for constant-returns-to-scale technologies, a very popular and theoretically defensible assumption in country growth studies. We use a simple time trend to model technical change but it should be noted that it can be directly proxied if one has country-specific observations on such factors as R&D expenditures, patent activity, etc., or if one uses the time dummy index approach of Baltagi and Griffin (1988).

The methods we consider in our empirical illustration of model averaging world productivity growth utilize both time invariant and time varying effects models to estimate innovation and efficiency change. There are many contributions to the efficiency and productivity literature that offer different ways to estimate this canonical panel model and to decompose TFP growth into a catch-up and innovation component. These include models introduced by Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), Lee and Schmidt (1993), Park, Sickles, and Simar (1998, 2003, 2007), Greene (2005a,b), and Kneip, Sickles, and Song (2012). Space limits the possibility of dealing with the many other approaches that have been proposed to estimate the panel stochastic frontier and provide a decomposition of TFP growth into innovative and catch-up, or technical efficiency. Bayesian estimators of panel stochastic frontiers have been developed by Liu, Sickles, and Tsionas (2017), which builds on earlier work by Van den Broeck, Koop, Osiewalski, and Steel (1994) and Tsionas (2006). The Bounded Inefficiency Model of Almanidis, Qian, and Sickles (2014) and related models of Lee (1996), Lee and Lee (2012), and Orea and Steinbuks (2012), and the "True" Fixed Effects Model of Greene (2005a,b) are also possible estimators to consider.

In order to keep the level of analysis workable but also to display the feasibility of utilizing a number of different methods and data specifications in our model averaging exercise, we limit our model to a workable set of 10 different models. These different estimators differ on the basis of what is assumed or not assumed about the form and structure of $\eta_i(t)$, of the idiosyncratic error v_{it} and in various independence assumptions on the possible correlations of $\eta_i(t)$ with the regressors x_{it}. The different estimators we used to construct innovation and efficiency change from the basic equation above for the basic Cornwell et al. (1990) efficient instrumental variables estimator (EIV) and time varying random effects estimator (CSSG), the Battese and Coelli (1992) time varying random effects estimator (BC), the Park, Sickles, and Simar (1996) semi-nonparametric efficient estimators (PSS1), the Park, Sickles, and Simar (1999) semi-nonparametric efficient estimator with serially correlated errors (PSS2W) and (PSS2G), standard fixed effects with no temporal variability in the effects (FIX1), standard random effects with no temporal variability in the effects (RND1), and fixed effects and random effects estimators with a quadratic and linear time trend for productivity growth (FIX2) and (RND2). Various assumptions on and specifications $\eta_i(t)$, v_{it}, the regressors x_{it}, and the correlation patterns among them are what differentiate these ten estimators. For example, the Cornwell et al. EIV estimator specifies $\eta_i(t)$ as a quadratic function of time for each cross-section and allows correlation between a subset of regressors and the time-varying effects $\eta_i(t)$. It is from estimates of the $\eta_i(t)$ that efficiency levels (renormalized and identified as technical efficiency) and changes in them over time (indicating efficiency growth) are constructed in the stochastic frontier literature. Technical change is estimated by including a time trend for all countries, which is analogous to constructing a Solow-type residual and relating it to a time trend. The EIV estimator is a Hausman and Taylor (1981) type estimator that utilizes orthogonality conditions (in this case between labor inputs and the time trend) and efficiency effects. The various semi-nonparametric panel estimators of Park, Simar, and Sickles are based on methods from the statistics literature that are used to develop estimators that are the most efficient within a class of estimators that are consistent with the parametric assumptions. These estimators utilize kernel smoothers to nonparametrically model correlation patterns and joint distributions for the effects and the possibly correlated regressors. The BC estimator assumes a parametric distribution for the effects and the error term and allows for time-varying effects that have the same patterns but can shift among the various countries. Testing for each of these specifications using standard specification metrics and then picking the winner introduces pre-test bias as well as assumes that there is one correct model. We argue below that such an approach is not warranted and that a consensus estimator based on weighting all of these models is.

We first detail the different models and then use them and the method of model averaging to evaluate world productivity trends from 1970 to 2000. In the Appendix A we also provide results from a limited Monte Carlo experiment to examine the finite sample properties of the model averaging methodology we employ in our empirical analyses.

5 Descriptions of Panel Data Stochastic Frontier Models

The generic production function that is estimated by all of the productivity estimators and then model averaged is the panel data stochastic frontier considered in Pitt and Lee (1981) and Schmidt and Sickles (1984), extended to address time-varying inefficiency by Cornwell, et al. (1990), which we introduced in the previous section. We first consider a special case of the general model above and write it as:

$$y_{it} = \alpha + X'_{it}\beta + \varepsilon_{it}, \quad i = 1, ...N, t = 1, ...T. \tag{4}$$

where

$$\varepsilon_{it} = v_{it} - u_{it}.$$

Here i indexes firms and t indexes time periods. y_{it} is the (log) output observation of firm i at tth period. X'_{it} is a vector of K (log) input observations of firm i at tth period. β are unknown parameters. v_{it} is distributed as $i.i.d$ $N(0, \sigma_v^2)$, u_{it} is a one-sided non-negative error and represents technical inefficiency. We next briefly explain each estimator we use in our model averaging analysis of world productivity growth.

5.1 Fixed Effects Estimator

The fixed effects estimator (FIX) is the fixed effect panel data model with time-invariant efficiency introduced by Schmidt and Sickles (1984). Here we assume $u_{it} = u_i$. Let $\alpha_i = \alpha - u_i$. Then the model becomes: $y_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it}$. The model then can be estimated following standard fix effect estimation.

There are the main advantage of this estimator is that the consistency of parameter estimates does not depend on the uncorrelatedness of the regressors and the individual effects. This is needed in the random effects specification considered by Pitt and Lee (1981). The slope estimator is consistent for either large N or T. However, consistency of the intercept α_i requires large T. An advantage of the fixed effects estimator is that consistency does not depend on the distribution of the effect as in Pitt and Lee (1981). Efficiency of the ith cross-section (country) is estimated as $\widehat{u}_i = \widehat{\alpha} - \widehat{\alpha}_i$ where $\widehat{\alpha} = \max(\widehat{\alpha}_i)$.

There is of course the issue of how we distinguish unobserved heterogeneity from inefficiency in the panel stochastic frontier model, a point raised by Greene (2005a,b) and explored by many in the last decade. In the panel stochastic frontier production function, output (or its log) for firm i at time t is a function of measures of inputs and observable variables to control for the production environment. Recall that the (log) linear production function (or distance function, recognizing that in the latter case x will contain right-hand-side endogenous multiple outputs) that we consider in this paper is $y_{it} = X'_{it}\beta + \eta_i(t) + v_{it}$. In order to make the point about what are or are not "true" fixed effects (of course this is just a semantic term of art as 'truth' in the context of unobserved heterogeneity is a misnomer) first decompose $\eta_i(t)$ into two terms, call them a_i and b_i. Differences across firms in the value of b_i reflect differences in the technical efficiency of production, and as in Schmidt and Sickles (1984) a conceptual measure of inefficiency can be expressed as

$$RIE_i = \max b_i - b_i \geq 0. \tag{5}$$

Differences in the value of a_i, however, reflect differences in the production environment that are beyond the control of the firm (or in our case, the country) and which we do not wish to include in our efficiency measures. Amsler and Schmidt (2016) use the following argument to make the point about how the interpretation of the "effects" as well as the modeling scenario are essential in decomposing and giving meaning to the a_i and b_i measures. Suppose that the firms are farms, in which case a natural interpretation of b_i would be a measure of the skill of the farmer. On the other hand, a_i would represent relevant but unobserved features of the production environment like soil quality or microclimate. It is clear that without additional assumptions, typically in the form of some type of orthogonality assumption, a_i and b_i cannot be separately point identified. The identification strategy of Amsler and Schmidt (2016) is to assume that there are some observable

variables that are correlated with a_i but not b_i and some other variables that are correlated with b_i but not a_i. Staying with their agricultural example, assume that the education of the farmer is correlated with her ability farmer but not with soil quality or microclimate. Define an indicator variable for the physical location of the farm that is correlated with soil quality or microclimate but not with the ability of the farmer. Then such a variables can be used as an instrument for b_i. Orthogonality conditions such as these, stated in terms of both of simple and partial correlations, are used by Amsler and Schmidt (2016) to develop identification conditions for the inefficiency and heterogeneity effects. For example, their first model assumes that the ability of the farmer is uncorrelated with physical location of the farm, whereas the second model assumes that, conditional on education of the farmer, ability of the farmer is uncorrelated with physical location of the farm. Using these assumptions various point identification results follow. There are of course many other considerations to explore in regard to decomposing the effects and imposing orthogonality conditions to identify them. As this paper is concerned largely with developing a methodology to construct a consensus estimate of TFP growth for policy makers based on model averaging methods we do not explore in this paper other potential sources of endogeneity not addressed by the various types of "within" estimators we utilize. Moreover, as we are not interested in the levels of efficiency or productivity but rather their changes, "fixed effects" *per se* are differenced away in our construction of productivity growth measures.

5.2   Random Effect Estimator

The random effects panel stochastic frontier estimator (RND) is a renormalized random effects estimator and was introduced by Pitt and Lee (1981). Their MLE estimator for such a model assumes that the time-invariant effect u_i is uncorrelated with the regressors and distributed as $i.i.d$ (μ, σ_u^2). The GLS estimator for the error components random effects stochastic frontier was considered by Schmidt and Sickles (1984). If one lets $\alpha^* = \alpha - \mu$, $u_i^* = u_i - \mu$, then a simple reparameterization of the model yields

$$y_{it} = \alpha^* + X_{it}'\beta + (v_{it} - u_i^*), \tag{6}$$

where the composed error has the standard form for the random components and feasible GLS can be employed. Given estimates of $\widehat{\beta}$, we can estimate u_i^* by $\frac{1}{T}\sum_t (y_{it} - \widehat{\alpha^*} - X_{it}'\widehat{\beta})$ and then estimate efficiency via $\widehat{u_i} = \max_i\{\widehat{u_i^*}\} - \widehat{u_i^*}$.

5.3   Cornwell-Schmidt-Sickles Estimator

Cornwell *el al.* (1990) (CSS) introduced a new panel data model with heterogeneity in both slopes and intercepts. The model allows them to estimate time-varying efficiency levels without imposing strong distributional assumptions for technical inefficiency or random noise.

   The model is written as:

$$y_{it} = X_{it}'\beta + Z_i'\gamma + W_{it}'\delta_i + \varepsilon_{it}, \qquad i = 1,...,N, t = 1,...T, \tag{7}$$
$$\delta_i = \delta_0 + u_i$$

We can rewrite the model as

$$y_{it} = X_{it}'\beta + Z_i'\gamma + W_{it}'\delta_0 + v_{it}, \tag{8}$$
$$v_{it} = W_{it}'u_i + \varepsilon_{it}$$

Here the u_i is distributed as $i.i.d$ $(0, \Delta)$, ε_{it} is distributed as $i.i.d$ $(0, \sigma^2)$, and is uncorrelated with regressors and u_i.

   In matrix form the model is:

$$y = X\beta + Z\gamma + W\delta_0 + v, \tag{9}$$
$$v = Qu + \varepsilon$$

where $Q = diag(W_i)$.

CSS within (hereinafter CSSW) does not assume that Qu is uncorrelated with the regressors. Defining the projection matrices $P_Q = Q(Q'Q)^{-1}Q'$ and $M_Q = I - P_Q$, we obtain CSSW as

$$\widehat{\beta}_{CSSW} = (X'M_Q X)^{-1}X'M_Q y. \tag{10}$$

CSS generalized least squares (hereinafter CSSG) assumes (X, Z, W) are uncorrelated with Qu, and can be written as

$$(\widehat{\beta}, \widehat{\gamma}, \widehat{\delta}_0)_{CSSG} = [(X, Z, W)'\Omega^{-1}(X, Z, W)]^{-1}(X, Z, W)'\Omega^{-1}y \tag{11}$$

where $\Omega = cov(v) = \sigma^2 I_{NT} + Q(I_N \otimes \Delta)Q'$. In order to relax the restriction of time-invariant efficiency, CSS assume that the intercept, time and time^2 are regressors whose coefficients exhibit cross-sectional heterogeneity and this leads to a simple quadratic in time specification for the effects that takes the form: $\alpha_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2$. Thus in the above model we have $W'_{it} = [1, t, t^2]$, $\delta'_i = [\theta_{i1}, \theta_{i2}, \theta_{i3}]$. Time-specific efficiencies can then be estimated as $\widehat{u}_{it} = \widehat{\alpha}_t - \widehat{\alpha}_{it}$ where $\widehat{\alpha}_t = \max_j(\widehat{\alpha}_{jt})$. CSS also provide an efficient instrumental variables estimator (hereinafter EIV), which is an extension of the Hausman and Taylor (1981) estimator. The full set of instruments are contained in

$$A = \Omega^{1/2}(M_Q X_2, X_1, Z_1, W_1) \tag{12}$$

where X_1, Z_1, W_1 are variables in X, Z, W that are uncorrelated with the effects, and X_2 are variables that are correlated with the effect. M_Q is orthogonal projection matrix of Q. The EIV estimator is given by

$$(\widehat{\beta}, \widehat{\gamma}, \widehat{\delta}_0)_{EIV} = [(X, Z, W)'P_A \Omega^{-1}(X, Z, W)]^{-1}(X, Z, W)'P_A \Omega^{-1}y. \tag{13}$$

Details on how one can consistently estimate Ω^{-1} can be found in Cornwell et al. (1990).

5.4 Battese-Coelli Estimator

Battese and Coelli (1992) (BC) introduced a fully parameterized maximum likelihood estimator. They define the technical efficiency for a given firm as an exponential function of time. Their model is specified as

$$Y_{it} = f(X_{it;}\beta)\exp(v_{it} - u_{it}) \tag{14}$$
$$\text{and } u_{it} = \eta_{it}u_i = \{\exp[-\eta(t - T)]\}u_i$$

where v_{it} is an i.i.d. $N(0, \sigma_v^2)$ and u_{it} is an i.i.d. non-negative truncated $N(\mu, \sigma^2)$. Notice that individual firm effect u_{it} can decrease, remain constant or increase as t increases, where $\eta > 0$, $\eta = 0$ or $\eta < 0$, respectively. $\eta = 0$ is the case where firm efficiency is time-invariant. Technical efficiency $TE_{it} = \exp(-u_{it})$ is based on conditional mean. Details are provided by Battese and Coelli (1992). The mean technical efficiency of firms at the tth period $TE_t = E[\exp(-\eta_t u_i)]$, where $\eta_t = \exp[-\eta(t - T)]$.

5.5 Semi-parametric Efficient Estimators

A class of estimators that achieve the semi-parametric efficiency bound (SPE) were introduced in a series of papers by Park and Simar (1994) and Park *et al.* (1998, 2003, and 2006). The estimators vary on how the basic model assumptions are modified to accommodate a particular issue of misspecification with the underlying efficiency model. We consider a number of SPE estimators that differ on the basis of assumed orthogonality of effects and regressors, temporal variation in the efficiency effects, and correlation structure of the population disturbance. The notion of efficient bounds in semi-parametric models has been well established in econometrics and statistics literature. The basic idea is to project the scores with respect to the slope parameters onto the nuisance parameter tangent space: $\pi(l_\theta | [l_\eta])$. Then we obtain efficient scores which are orthogonal to the

9

scores of nuisance parameters: $l^* = l_\theta - \pi(l_\theta|[l_\eta])$. Thus we can obtain Fisher information bound $E(l^* l^{*-1})$. For details, see Newey (1990).

Park *et al.* (1998) explore the semi-parametric efficient estimation of stochastic frontier models in which the effects and the regressors have certain dependency structures. They discuss three time invariant models. The first model assumes no particular structure of dependence between the effects and the regressors, which is analogous to the fixed effect estimator. The second model assumes dependency between the effects and a subset of regressors, which is analogous to the Hausman and Taylor estimator. The third model (PSS1) allows for dependency between the effects and long run movements in a subset of regressors. They derive semi-parametric efficiency bound for each model, and methods to estimate parameters and effects.

Park *et al.* (2003) focus on the semi-parametric efficient estimation of random effect panel models containing AR(1) disturbances:

$$Y_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it}, \, i = 1, ..., N, t = 1, ..., T, \tag{15}$$
$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + u_{it}, |\rho| < 1$$

and u_{it}s are distributed as i.i.d. $N(0, \sigma^2)$. Denoting $X_i = (X'_{i1}, ..., X'_{iT})'$, (α_i, X_i) are independent of ε_i and are i.i.d. random variables having unknown density $q(.,.)$ on R^{1+dT}. They consider two structures describing the relationship between X and α: Model 1 (PSS2G) assumes the independence between X_i and α_i; Model 2 (PSS2W) allows dependence between X_i and α_i. They then provide semi-parametric efficiency bound for each model, and methods to estimate parameters and effects.

Park *et al.* (2007) extend the semi-parametric efficient estimation to dynamic panel data models (PSS3). Their model can be written as:

$$Y_{it} = \gamma Y_{i,t-1} + X'_{it}\beta + \alpha_i + \varepsilon_{it}, \, i = 1, ..., N, t = 1, ..., T, \tag{16}$$

where ε_{it}s are distributed as i.i.d. $N(0, \sigma^2)$. Based on assumptions regarding conditional independence, their models use non-parametric estimators for the random effects, and use parametric assumptions on the distribution of the within errors.

A summary of the estimators we consider in our model averaging analysis is presented in Table 1 below.

	Table 1	
Estimator	Reference
FIX	Schmidt and Sickles (1984)
RND	Schmidt and Sickles (1984)
CSSG, EIV	Cornwell, Schmidt and Sickles (1990)
BC	Battese and Coelli (1992)
PSS1	Park, Sickles and Simar (1998)
PSS2W, PSS2G	Park, Sickles and Simar (2003)
PSS3	Sickles (2005), Park, Sickles and Simar (2007)

6 Discussion on Combining Estimates

Before we move to our evaluation of the model averaging results we want to focus a bit of attention on why such approaches are both statistically reasonable and intuitive. We first note that combining estimates from different models provides a solution to modeling uncertainty. Sickles (2005) pursued this strategy in his examination of semiparametric and nonparametric panel frontier estimators. The literature on the topic of model averaging is deep and interest in it as has accelerated as "big data" and issues related to its uses have become mainstream topics in academic research and private sector decision-making. Hjorth (1994), Burnham and Anderson (2002), and Leeb and Potscher (2005) all made important early contributions. A theme that links all of these early and subsequent contributions is that procedures to approximate all possible data generating processes be considered.

This is a natural implication to George Box's famous quote that, "*essentially, all models are wrong, but some are useful*" (Box and Draper, 1987, p. 424).

There is also a variety of justifications for considering such alternative explanations that come from theoretical perspectives in economics. Work from the social choice literature suggests that the majority voting outcome will be the median candidate (Moulin, 1980). If voter preferences are symmetric then the median is equal to the mean. The mean is a particular aggregator function. Social choice theory also speaks to decision rules wherein expected outcomes are based on an aggregator function with weights based on the R^2 associated with the estimates of each expected outcome. This is the Tullock (1980) *contest function*.

Insights from statistics come from *model averaging* and methods of *combining forecasts*. In order to implement model averaging one must assign weights to each possible set of model estimates and this leads to weights based on the Akaike Information Criterion (Akaike, 1973), hereinafter AIC), Mallows' CP (Mallows, 1973), and the Bayesian Information Criterion (Schwarz, 1978) (BIC). Buckland, et al. (1997) used two of these information criteria, the Akaike and Schwarz, as weights in their model averaging exercise. Hansen (2007) showed that the Mallows' Model Average estimator is asymptotically optimal in some cases and more favorable compared to AIC and BIC. Carroll, Midthune, Freedman, and Kipnis (2006) conducted a nutritional epidemiologic study and showed AIC achieved an efficiency gain, whereas BIC had serious issues and was not recommended. Burnham and Anderson (2002) and Claeskens and Hjort (2008) have more detailed discussions of the literature. Hansen and Racine (2012) considered situations in which candidate models are non-nested proposed a jackknife model averaging estimator, which they showed is asymptotically optimal in the sense that it approaches the lowest possible expected squared errors. Simulated comparisons of criteria have also been studied in different disciplines. Parmeter, Wan, and Zhang (2015) have begun to assess the finite sample properties of this estimator via Monte Carlo simulations.

The statistical literature involving *combining* time-series *forecast* models also provides us with justifications for the approaches we take to develop consensus TFP estimates. Important contributions to this literature include Bates and Granger (1969), Lahiri, Teigland, and Zaporowski, (1988), Zarnowitz and Lambros (1992), Davies and Lahiri (1995), Diebold and Lopez (1996), Newbold and Harvey (2002), Lahiri and Sheng (2010), and Lahiri, Peng and Zhao (2017).

The weights we use in our model averaging exercise are based on those discussed in Buckland et al. (1997), which proposes a frequentist model averaging methodology motivated by Bayesian Model Averaging, which is itself based on Bayesian model selection criteria. They propose assigning weights according to the information criteria

$$I_m = -s \log(L_{m)} + l, \tag{17}$$

where L_m is the maximized likelihood function of the m-th model, and l is some penalty function. Buckland et al. (1997) note that if one assumes that the penalty for the different models is comparable and the set of models under consideration constitute the finite set of possible models then, based on the Schwarz (1978) approximation of the Bayes factor, a plausible choice for the weight is:

$$w_I = \frac{\exp(\frac{1}{2}I_m)}{\sum_{m \in \mathcal{M}} \exp(\frac{1}{2}I_m)} \tag{18}$$

in which $\mathcal{M}$ is the set of all competing models. This weighting algorithm nests all of the various methods we utilize in our analyses. For example, if $l = 2k$, where k is the number of parameters in the model, then the information criterion I_m is just the AIC score. If $l = k \cdot \log(NT)$, where NT is the number of panel observations, I_m becomes the BIC score. When we use regression-based approaches in our analysis we utilize the standard measures of the AIC and BIC (also refereed to as the Schwarz criterion or SC) model selection criteria in terms of the regression residuals based on the following equalities

$$AIC = \ln(\frac{SSE}{NT}) + \frac{2k}{NT} \tag{19}$$

$$BIC \ (SC) = \ln(\frac{SSE}{NT}) + \frac{k \ln(NT)}{NT}. \tag{20}$$

Weights based on the other measures of model fit that we utilize, the residual sum of squares (RSS), the (adjusted) R^2 and simple arithmetic averages, use various reparameterizations of these expressions.[1]

7 Modeling World Economic Growth with the UNIDO Data

In order to better understand existing patterns of world income levels, growth in per/capita income, political stability, and international trade flows, and to reasonably anticipate future trends in these important economic indicators, it is important to correctly measure countries' productivity growth. In addition, when gauging crucial economic indicators such as these it is important to use methods which are robust to misspecification error. The following section deals with productivity growth measurement's robustness using a number of economic methodologies, and estimators consistent with them, to elucidate productivity growth. Our model averaging productivity procedures use the United Nations Industrial Development Organization (UNIDO) data over the period between 1970 and 2000. The issue of country heterogeneity is addressed in part by separately analyzing countries grouped by their development features and by using different panel data methods to address cross-section and time-specific unobserved effects.

7.1 UNIDO Data Description

Information on measures of the level and growth of TFP based on 12 different empirical approaches across 112 countries over the period $1960-2000$ are provided by The World Productivity Database (WPD). The Penn World Tables version 6.1 (PWT, Heston, Summers and Aten, 2002) is the primary source of data. The Penn World Tables version 6.1 was used to obtain (chain weighted) GDP and investment in power purchasing parity 1996 US dollars. Data on employment and hours worked were taken from the Groningen Growth and Development Centre (GGDC, 2005) and Asian Development Bank (ADB, various issues). The International Labor Organization (ILO) Yearbook 2003 was the source of unemployment figures, schooling data was taken from the ILO's Key Indicators of the Labour Market and ADB (various issues), Barro and Lee (2000) and the health indicators -life expectancy and adult mortality rates- were obtained from the World Development Indicators (World Bank, 2004). These are documented in the technical appendix to the WPD (Isaksson, 2007)

Capital input measurement is the most complicated. It can be argued that Capital (K) is the hardest feature to measure, which is why the WPD has made 4 different approaches available based on various computations for the initial capital stock, depreciation rates, depreciation schedules, and the lifetime of the asset. The various capital measures are labelled K06, K13, and Keff. K06 and K13 are built on the assumption that ten years of investment act as a sufficient proxy for the initial capital stock K0. The two capital stocks only differ in terms of their assumed constant rates of depreciation, which are 6.0 and 13.3 percent, respectively (hence, K06 and K13). Another method of gauging capital examines the profile of capital productivity and uses a time-varying depreciation rate. The productive efficiency of the asset falls as the asset ages which results in Keff. Detailes of its computation can be found in Isaksson (2007).

Two types of labor utilization rates for which labor force can be adjusted are involved in *Labor input* measurement: differences in employee numbers and working hours. Employment (EMP) is the first measure for labor force (LF) and is a direct measure of employment. The second is achieved by using unemployment rates on LF data, resulting in derived employment (DEMP). Unfortunately missing observations for many countries for many of the years of our analyses forced us to utilize only LF as the labor input in our model averaging exercise.

---

[1]Clearly these last two measures of model fit do not penalize the weights for over-parameterization but as NT is rather large in our analyses such penalty terms are not as influential as they might otherwise be in studies with relatively smaller degrees of freedom. Such a conclusion is borne out by the similarity in the model averaged estimates based on different weights.

7.2 Empirical Findings

We adopted the approach taken by Hulten and Isaksson (2007) who divided all 112 countries in the WPD into six mutually exclusive groups, in accordance with the World Bank classification by income per capital. The group of Low Income countries (hereinafter LOW) is comprised of 40 countries, 22 countries in the group of Lower-Middle Income countries (hereinafter LOW-MID), 17 countries in the Upper-Middle Income countries (hereinafter UPPER-MID), 24 High-Income countries (hereinafter HIGH), 4 Old Tigers (the original Asian Four Tigers) and 5 New Tigers. The countries are:

Low Income Countries: Angola, Bangladesh, Benin, Bolivia, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Congo, Cote d'Ivoire, Democratic Republic of the Congo, Ethiopia, Gambia, Ghana, Guinea, Guinea Bissau, Haiti, Honduras, Kenya, Lesotho, Madagascar, Malawi, Mali, Mauritania, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Papua New Guinea, Rwanda, Senegal, Sierra Leone, Tanzania, Togo, Uganda, Zambia, and Zimbabwe.

Lower-Middle Countries: Algeria, Cape Verde, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Fiji, Guatemala, Guyana, Iran, Jamaica, Jordan, Morocco, Namibia, Pakistan, Paraguay, Peru, Philippines, and Sri Lanka.

Upper-Middle Countries: Argentina, Barbados, Botswana, Brazil, Chile, Gabon, Mauritius, Mexico, Panama, Seychelles, South Africa, Syria, Trinidad and Tobago, Tunisia, Turkey, Uruguay, and Venezuela.

High-Income Countries: Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Greece, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK, and USA.

Old Tigers: Hong Kong (SAR of China), Republic of Korea, Singapore, and Taiwan (Province of China).

New Tigers: China, India, Indonesia, Malaysia, and Thailand.

As we mention above, as a result of limitations in the availability of data, we used LF as the labor input. K06, K13, and Keff are the different capital input measures. Thus, every country group contains 3 input combinations. Ten different estimating models were used: EIV, CSSG, BC, PSS1, PSS2W, PSS2G, FIX1, RND1, FIX2 and RND2. The first six estimators use the methods discussed in Section 5 to estimate efficiency growth and innovation change, with the latter identified by a time trend that identifies a common technical change shared by each country group. The last four are simply fixed and random effects pooled estimates using a time trend or a time trend and a quadratic time trend to estimated average growth rates of TFP within the country group. This results in 30 different models to consider in our model averaging exercise. The period of observation is 1970 to 2000.

In our analyses, TFP growth is based on the contributions of the growth in technical efficiency and the growth in technical innovation. Although patterns of technical efficiency changes are revealing the level of these changes is relatively small compared to the level of technical change. LOW had substantial efficiency enhancements from 1970 to the early 1980s. LOW-MID had efficiency change patterns similar to LOW, but efficiency growth turns slightly negative at the turn of the 1980s. UPPER-MID had slight efficiency improvements through the middle of the 1990s. HIGH had minor growth in efficiency development through the 1980s. Efficiency growth for the Old Tigers showed a small reduction in efficiency at the beginning of the sample period, whilst a pattern of increasing efficiency was visible towards the end of the sample period. The efficiency trends for the New Tigers is akin to LOW, but are smaller in magnitude. More detailed results and figures for the efficiency results are available from the authors. We do not pursue the efficiency results in any greater detail as they have only a marginal impact on TFP growth in our regressions. Since TFP growth is dominated by technical change we now turn to the summary TFP results.

The Old Tigers have the highest TFP improvements, followed by New Tigers. HIGH and UPPER-MID income nations have modest TFP annual growth rates. LOW-MID and LOW income nations display relatively stagnant TFP growth. Projected accumulated TFP growth between 1970 and 2000 is 15.4% for LOW income nations, 10.1% for LOW-MID income nations, 27.7% for UPPER-MID income nations, 17.2% for HIGH income nations, 199.7% for the New Tigers, and 239.4% for the Old Tigers. We also compared our model averaged TFP results (using GDP share weights) with

results presented on the UNIDO WPD website. Five additional approaches were selected baselines against which to match our findings. These are labelled as Growth Accounting, Harrod Neutral, Pooled Regression, Panel Regression, Stochastic Frontier Model, and Data Envelopment Analysis (DEA) and represent standard approaches to measuring TFP growth used in many national and international agencies. They are calculated from the UNIDO website's online software packages found at the UNIDO website (https://www.unido.org/data1/wpd/Index.cfm) and documented in Isaksson (2007). We selected identical nations as well as identical input groupings for comparability to our model averaged estimates. The Growth Accounting estimate of TFP growth is based on the standard discrete growth equation $\Delta \ln A_t = \Delta \ln Y_t - \alpha^* \Delta \ln K_t - \beta^* \Delta \ln L_t$, where income shares are assumed to be constant, both in time and in space within the various country groupings. The Harrod neutral TFP estimates allow for the production function to shift along a constant capital-output ratio, instead of a constant capital-labor ratio and thus not only involves a shift, but also a tilt of the production function, which changes the marginal product of capital. The growth equation is thus modified to be $\Delta \ln A_t = \Delta \ln Y_t - \frac{\alpha}{\beta}^*(\Delta \ln \kappa_t - \Delta \ln \kappa_{t-1})$, where $\kappa = K/Y$. The Pooled Regression estimates are based on a simple regression for the Cobb–Douglas specification with a time trend to account for TFP change and the pooling is over the particular country group. Panel Regression results are based on a fixed effects specification of the country effects with a time trend. We also use this specification as one of the competing models in constructing our model averaging estimates. Stochastic Frontier Model estimates are based on the random-effects model of Battese and Coelli (1992), which we also use in our model averaging estimator. Finally, the Data Envelopment Analysis estimates are based on the standard linear programming methods outlined in Färe, et al. (1994), Coelli, et al. (1998), and discussed in the context of the UNIDO data in Isaksson (2007).

Figures $1-6$ display TFP growth rates for nations in the various developmental classifications. Our model averaging results of course indicate smoother patterns as a result of the averaging. For four out of the six sets of countries in different levels of development our model averaged TFP growth rates surpassed the ones projected from the five approaches used on the UNIDO WPD website. These ranged from a high of about 3.6% annual TFP growth for the Old Tigers to a low of about 0.32% annual growth for the Low-Mid income countries.

The last results we discuss are the combined estimates and their distribution within the differing country groupings we adopted for our analysis. Again, the motivation for employing the model averaging methodology is to obtain consensus results based on all the modeling and data information at hand. The annual changes of technical efficiency, technical innovation and TFP for each individual group based on different weighting methods are shown in Table 1. The most crucial component the model averaging exercise is the assignment of weights to each set of estimates. The simplest averaging is to take the arithmetic mean of all estimates, which implicitly assumes equal importance of all models. Besides simple averaging, we use other four statistical criteria that have been developed to measure model fit and to assess model specification in order to assign weights. These are based n the methods we detailed above and based on the model selection criteria: (adjusted) R^2, residual sum of squares (RSS), Akaike's information criterion (AIC), Bayes Information Criteria (BIC), and simple arithmetic averages. These different weighting approaches are used to develop different model averaged estimates for annual changes of technical efficiency, technical innovation and TFP by weighting criteria. For example, using the simple arithmetic averaged, the model averaged estimates for annual changes of technical efficiency, technical innovation and TFP are $0.64\%, -0.17\%$ and 0.47% for LOW, $0.35\%, -0.04\%$ and 0.32% for MID-LOW, $0.15\%, 0.64\%$ and 0.79% for UPPER-MID, $-0.06\%, 0.57\%$ and 0.51% for HIGH, $0.30\%, 3.33\%$ and 3.63% for the Old Tigers, and $-0.09\%, 2.95\%$ and 2.86% for the New Tigers. Since (adjusted) $R^{2\prime}s$ for our different specifications are rather close to each other, the weighted results are quite similar using the other criteria as they all are comparable when degrees of freedom are relatively large, which they are for the relatively parsimonious Cobb-Douglas production function-based productivity estimates. The combined estimates display a relatively stable pattern during the 31 years covered in the sample period as Figure 7 illustrates. Old Tigers lead the world in TFP improvement due mainly to substantial technological innovations with technical efficiency change having a small positive contribution. New Tigers were next with quite respectable TFP growth to innovation change with no economically significant contribution by technical efficiency change. LOW income countries had

14

the greatest efficiency gains among all groups but their poor innovation growth leads to poor overall TFP performance. LOW-MID and UPPER-MID countries have essential no progress in either technical efficiency or innovation. HIGH income countries face little growth in innovation as well as a slight regress in their levels of technical efficiency.

Bates and Granger (1969) introduce the methodology of forecast combination. In their paper, clearly the results they attempt to combine are correlated since the outcomes are obtained by two different forecast methods but on the same data set. In their first weighting method, if the forecast errors σ_1^2, σ_2^2 from the two models are uncorrelated, to minimize the total error, the weights should be assigned as $\sigma_2^2/(\sigma_1^2 + \sigma_2^2)$ and $\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. The weighting will be a little bit more complicated if correlation is considered: weight for forecast 1 will be $(\sigma_2^2 - \rho\sigma_1\sigma_2)/(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$. If the weights are decided as above, the variance of the forecast error is no greater than the smaller values of the two variances. It is obvious that the bigger error variance result will receive smaller weights. If only two results are combined, the weights trivially are the same as one of our weighting criteria which we assign $1/\sigma_1^2$ to estimate 1 and $1/\sigma_1^2$ to estimate 2. The method they applied to forecast model can be used in our study since it is to minimize combined error, whether it is an out-of-sample forecast error or in-sample error. Generally speaking, all the weight selecting methods are based on some types of loss function which in turn rely on the differences between the realized outcome and the forecast outcome, such as a Mean Squared Error (hereinafter MSE) or Mean Squared Percent Error (hereinafter MSPE). If we choose the loss function as typical square of error, it would be perfectly reasonable to use "goodness-of-fit" criteria. The implementation will be identical to the model averaging we discussed earlier. Developments in choosing weights by applying automatic machine learning algorithms and methods to solve missing data problems have been discussed in Lahiri, Peng and Zhao (2017).

For inference purpose, the variances of combined estimates can be calculated under model averaging framework. Burnham and Anderson (2002) and Huang and Lai (2012), have provided discussions on how to compute them. The difficult component to estimate is the correlations between each pair of estimators. For example, in our case it is the estimate of TFP from each statistical model based on a particular combination of variously defined capital and labor inputs. Bootstrap methods have been suggested in these referenced studies. However, bootstrapping methods are difficult to implement given the various forms of unobserved heterogeneity and potential error correlations. In the situations when correlations cannot be estimated, an upper bound on variance can still be obtained assuming all correlations are 1. However, estimating sample correlations between each pair of estimates is not difficult in our study because of our panel-data setting. In the majority of our models TFP estimates are time-varying. Thus we can calculate sample correlations between each pair of TFP directly because we have estimates in each period. Such estimates are not possible for models with time-invariant TFP estimates and for those we use the upper bound on the variance of 1. The associated variances and variance bounds for our model averaged TFP estimates are presented on Table 1. The variances of the combined estimates can also be calculated under the model averaging framework and thus the statistical significance of the consensus estimates can also be determined. In this way such consensus estimates provide an advantage over the index number measures that are usually presented simply as a point estimate.

Badunenko et al. (2013) and Henderson and Russell (2005) have utilized programming methods and data envelopment analysis (DEA) to study efficiency growth using a somewhat different set of country groupings. Their DEA analyses attribute all measurement error to the one-sided inefficiency component as opposed to the stochastic frontier methods we utilize that allows for measurement error. We can highlight similarities and differences in our study theirs based on the alternative DEA methodology[2]. Between 1965 and 2000, they report total productivity (output per worker) increases of 114.8% (annualized at 2.10%) in Asia (notice their included countries in this region are significantly different from ours), 26.6% (annualized at 0.66%) in Latin America, and 110.2% (annualized at 2.10%) in the OECD. Even though the two studies employ different methodologies on different data sets, they both find that countries in Asia have the highest TFP growth, those

---

[2]In Asian countries, they have India, Indonesia, Iran, Jordan, Malaysia, Nepal, Syria and Thailand. In Latin America, they have all the 11 countries we have in addition to 9 other countries. In OECD countries, Iceland and Luxembourg are not included in their data set, Mexico is not included in our data set.

that constitute the OECD have the second highest TFP growth rates, and countries in Latin (and South) America have the lowest TFP growth rates. Another common finding is that technical innovation contributes significantly more than the efficiency gains to TFP growth in Asia and the OECD. That is, shifts in the production possibility frontier outweigh the catch-up effect in the generation of TFP growth. Since TFP growth is much smaller than the growth of GDP per capita through the sample period, other factors such as such as physical and human capital accumulation play a relatively larger role in world economic growth.

8 Conclusion

In this paper we have focused on the role that panel data econometrics plays in formulating and estimating the most important contributors to productivity growth: innovation and catch-up. We have explained different theories on economic growth and productivity measurement and the econometric specifications they imply. Various index number and regression-based approaches to measuring productivity growth and its innovation and catch-up components have been discussed in detail. We have also discussed methods that can be used to combine results from the many different perspectives on how economic growth is modelled and estimated, focusing on methods used in model averaging and in the combination of forecasts. We have utilized various panel data and model averaging methods in an analysis of world productivity growth using the WPD and have analyzed as well the changes in the US income distribution that would have resulted by relinking the growth of TFP with the growth in wage compensation, which were connected so strongly in the post-WWII to early 1970's era.

The motivation for employing the model averaging methodology is to arrive at a consensus result based on all the modeling and data information at hand. To this end, we started by creating country groups based on income levels (LOW, LOWER-MID, UPPER-MID and HIGH), however singling out two fast-growing groups of developing countries-Old and New Tigers-producing a total of six groups of countries. Based on results from ten estimating models and three different capital definitions (a total of 30 different sets of estimates), we found that Old Tigers have the highest TFP improvements thanks to the group's relatively large level of technical progress. New Tigers were the second best performer, despite some small deterioration in technical efficiency over the sample period. Both HIGH and UPPER-MID display negative trends in TFP growth, but still manage moderately positive TFP growths every year. The worst outcome is shown for LOW-MID and LOW because of lack of technological innovation and decline in technical efficiency. When aggregating these results to obtain a world average we found a declining trend in TFP growth rates.

In a second step, we compared our average estimate with those obtained from growth accounting (Hicks-neutral), Pooled Regression, Panel Regression, Stochastic Frontier Model and Data Envelopment Analysis. Compared to these models, our estimates display smoother trends, with four of the ten estimating models showing higher TFP growth than the five comparators. It is reassuring that our results are robust to different approaches to weighing each set of estimates, which is the most crucial component for combining estimates.

While our paper is an important step in the direction of obtaining a consensus TFP result much work remains. For example, in this paper we have used ten estimating models but why not use more models? Secondly, in this paper we have not experimented much with other sources of TFP growth variation such as additional production factors (e.g., human capital, health capital and land) or functional forms (e.g., Cobb-Douglas or CES). Given that different countries are at various stages of development, applying different properties of the production function might be an important future step; at least it would be good to see an attempt being made in that direction. Thirdly, there may be scope for experimenting with country groupings based on other criteria than income and level of development alone. For example, it is likely that countries whose economies are highly leveraged on natural resources, such as the Gulf States or countries in transition (Eastern Europe), might display alternative performance patterns. Similarly, as more country data and longer time series become available, opportunities for even richer and nuanced analysis may arise.

9 References

Abramovitz, M. (1986), "Catching Up, Forging Ahead, and Falling Behind,"*Journal of Economic History,* 46, 385–406.

Ahn, S. C., Good, D. H., and R. C. Sickles (2000), "Estimation of Long-run Inefficiency Levels: A Dynamic Frontier Approach," *Econometric Reviews,* 19, 461–492.

Aigner, D., Lovell, C. A. K., and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* 6, 21–37.

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Csaki, eds., *Second International Symposium on Information Theory,* 267-281, Akademiai Kiadó, Budapest.

Almanidis, P., Qian, J. and R. C. Sickles (2014), "Stochastic Frontiers with Bounded Inefficiency," *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications,* R. C. Sickles and W. C. Horrace, eds., 47-82, Springer Science & Business Media, New York.

Amsler, C. and Schmidt, P. (2016). Distinguishing heterogeneity and inefficiency in a panel data stochastic frontier model. Michigan State University, mimeo.

Arrow, K. J. (1962), "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29, 155–173.

Bai, J., Kao, C. and S. Ng (2009), "Panel Cointegration with Global Stochastic Trends," *Journal of Econometrics*, 149, 82–99.

Baltagi, B. H. and J. M. Griffin (1988), "A General Index of Technical Change," *Journal of Political Economy*, 96, 20–41.

Bates, J. M. and C. W. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly,* 20, 451–468.

Badunenko, O., Henderson, D. J. and R. R. Russell (2013), "Polarization of the Worldwide Distribution of Productivity," *Journal of Productivity Analysis*, 40, 153–171.

Blazek, D. and R. C. Sickles (2009), "Knowledge and Geographical Spillovers in Wartime Shipbuilding," *Economic Modeling*, 27, 1484–1497.

Box, G. P. and N. R. Draper (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.

Buckland, S. T., Burnham, K. P., and N. H. Augustin (1997), "Model Selection: An Integral Part of Inference," *Biometrics*, 53, 603-618.

Burnham, K. P. and D. R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer Verlag, New York.

Carroll, R. J., Midthune, D., Freedman, L. S. and V. Kipnis (2006), "Seemingly Unrelated Measurement Error Models, With Application to Nutritional Epidemiology," *Biometrics*, 62, 75–84.

Claeskens, G. and N. L. Hjort (2008), *Model Selection and Model Averaging (Cambridge Series in Statistical and Probabilistic Mathematics)*,Cambridge University Press, New York.

Coe, D. T. and E. Helpman (1995), "International R&D Spillovers," *European Economic Review*, 39, 859–887.

Coelli, T.J., D.S.P. Rao, and G.E. Battese (1998), *An Introduction to Efficiency and Productivity Analysis*, Boston: Kluwer Academic Publishers.

Cornwell, C., Schmidt, P. and R. C. Sickles (1990), "Production Frontiers with Cross-sectional and Time-series Variation in Efficiency Levels," *Journal of Econometrics*, 46, 185–200.

Cuesta, R. A. (2000), "A Production Model With Firm-Specific Temporal Variation in Technical Inefficiency: With Application to Spanish Dairy Farms," *Journal of Productivity Analysis*,13, 139–158.

Davies, A. and K. Lahiri (1995), "A New Framework for Analyzing Survey Forecasts using Three-dimensional Panel Data," *Journal of Econometrics*, 68, 205–227.

Diao, X., Rattsø, J. and H. E. Stokke (2005), " International Spillovers, Productivity Growth and Openness in Thailand: An Intertemporal General Equilibrium Analysis," *Journal of Development Economics*,76, 429–450.

Diebold, F. X. and J. A. Lopez (1996), "Forecast Evaluation and Combination," *National Bureau of Economic Research Working Paper No. 192*, Cambridge, MA.

Dowrick, S. and D.-T. Nguyen (1989), "OECD Comparative Economic Growth 1950-85: Catch-up and Convergence," *American Economic Review,* 79, 1010–1030.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994), "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *American Economic Review*, 84, 66-83.

Good, D. H., Nadiri, M. I. and R. C. Sickles (1997), "Index Number and Factor Demand Approaches to the Estimation of Productivity," in M. H. Pesaran and P. Schmidt, eds., *Handbook of Applied Economics, Volume II-Microeconometrics*, Basil Blackwell,Oxford, 14-80.

Greene, W. (2005a), "Fixed and Random Effects in Stochastic Frontier Models," *Journal of Productivity Analysis*, 23, 7–32.

Greene, W. (2005b), "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model," *Journal of Econometrics,* 126, 269–303.

Groningen Growth and Development Centre (2005), *GGDC Data Bases,* University of Groningen, Netherlands.

Grosskopf, S. and S. Self (2006), "Factor Accumulation or TFP? A Reassessment of Growth in Southeast Asia," *Pacific Economic Review,* 11, 39–58.

Hansen, B. E. (2007), "Least Squares Model Averaging," *Econometrica,* 75, 1175–1189.

Hansen, B. E. and J. S. Racine (2012), "Jackknife Model Averaging," *Journal of Econometrics*, 167, 38-46.

Hausman, J. A., and W. E. Taylor (1981) "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

Henderson, D.J. and R. R., Russell (2005), "Human Capital and Convergence: A Production-Frontier Approach," *International Economic Review*, 46, 1167-1205.

Heston A., Summers R. and B. Aten (2002), *Penn World Table Version 6.1*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, Philadelphia.

Hjorth, J. U. (1994), *Computer Intensive Statistical Methods: Validation Model Selection and Bootstrap*, Chapman & Hall/CRC, Boca Raton, FL.

Huang, C. J. and H.-P. Lai (2012), "Estimation of Stochastic Frontier Models Based on Multimodel Inference," *Journal of Productivity Analysis,* 38, 273–284.

Hultberg, P. T., Nadiri, M. I., and R. C. Sickles (1999), "An International. Comparison of Technology. Adoption and Efficiency: A Dynamic. Panel Model," *Annales d'Economie et de Statistique,* 55-56, 449–474.

Hultberg, P. T., Nadiri, M. I. and R. C. Sickles (2004), "Cross-country Catch-up in the Manufacturing Sector: Impacts of Heterogeneity On Convergence and Technology Adoption," *Empirical Economics*, 29, 753–768.

Hulten C. R. and A. Isaksson A. (2007), "Why Development Levels Differ: The Sources of Differential Economic Growth in a Panel of High and Low Income Countries," *National Bureau of Economic Research Working Paper No. 13469*, Cambridge, MA.

Isaksson, A. (2007), "World Productivity Database: A Technical Description," *Research and Statistics Staff Working Paper No. 10*, United Nations Industrial Development Organization, Vienna.

Jorgenson, D. and Z. Griliches (1972), "Issues in Growth Accounting: A Reply to Edward F. Denison," *Survey of Current Business*, 52, 65-94.

Kim, J.-I. and L. J. Lau (1994), "The Sources of Economic Growth of the East Asian Newly Industrialized Countries," *Journal of the Japanese and International Economies*, 8, 235–271.

Kneip, A., Sickles, R. and W.-H. Song (2012), "A New Panel Data Treatment for Heterogeneity in Time Trends," *Econometric Theory*, 28, 590-628.

Krugman, P. (1994), "The Myth of Asia's Miracle," *Foreign Affairs* Nov/Dec, 62–78.

Kumbhakar, S. C. (1990), "Production Frontiers, Panel Data, and Time-varying Technical Inefficiency," *Journal of Econometrics*, 46, 201–211.

Lahiri, K., Peng, H. and Y. Zhao (2017), "On-line Learning and Forecast Combination in Unbalanced Panels," Econometric Reviews, 36, 257-288

Lahiri, K. and X. Sheng (2010), "Measuring Forecast Uncertainty by Disagreement: The Missing Link," *Journal of Applied Econometrics*, 25, 514–538.

Lahiri, K., Teigland, C. and M. Zaporowski (1988), "Interest Rates and the Subjective Probability Distribution of Inflation Forecasts," *Journal of Money, Credit and Banking*, 20, 233–248.

Lee, Y. H. (1996), "Tail Truncated Stochastic Frontier Models," *Journal of Economic Theory and Econometrics*, 2, 137–152.

Lee, Y. H. and S. Lee (2012), "Stochastic Frontier Models with Threshold Efficiency," *Journal of Productivity Analysis*, 42, 45-54.

Lee, Y. H. and P. Schmidt (1993), "A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency," in H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York, 237–255.

Leeb, H. and B. M. Pötscher (2005), "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

Liu, J., Sickles, R. C., and E. Tsionas (2017), "Bayesian Treatment to Panel Data Models with Time-varying Heterogeneity," Rice University, Houston.

Lucas Jr, R. E. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

Mallows, C. L. (1973), "Some Comments on Cp," *Technometrics*, 15, 661–675.

Meeusen, W. and J. van Den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review*, 18, 435–444.

Moulin, H. (1980), "On Strategy-proofness and Single Peakedness," *Public Choice*, 35, 437–455.

Nelson, R. R. and G. Wright (1992), "The Rise and Fall of American Technological Leadership: The Postwar Era in Historical Perspective," *Journal of Economic Literature*, 30, 1931–1964.

Newbold, P. and D. I. Harvey (2002), "Forecast Combination and Encompassing," in M. P. Clements and D. F. Hendry, eds., *A Companion to Economic Forecasting*, John Wiley & Sons, New York, 268–283.

Orea, L. and J. Steinbuks (2012), "Estimating Market Power in Homogenous Product Markets Using a Composed Error Model: Application to the California Electricity Market," *Cambridge Working Paper in Economics 1220, EPRG Working Paper 1210*, University of Cambridge.

Park, B. U., Sickles, R. C. and L. Simar (1998), "Stochastic Panel Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273–301.

Park, B. U., Sickles, R. C. and L. Simar (2003), "Semiparametric-efficient Estimation of AR (1) Panel Data Models," *Journal of Econometrics*, 117, 279–309.

Park, B. U., Sickles, R. C. and L. Simar (2007), "Semiparametric-efficient Estimation of Dynamic Panel Data Models," *Journal of Econometrics*, 136, 281–301.

Parmeter, C., Wan, A., and X. Zhang (2015), "A Model Averaging Stochastic Frontier Estimator," University of Miami, Coral Gables, Florida.

Pitt, M. and L.-F. Lee (1981), "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry," *Journal of Development Economics*, 9, 43–64.

Reikard, G. (2005), "Endogenous Technical Advance and the Stochastic Trend in Output: A Neoclassical Approach," *Research Policy*, 34, 1476–1490.

Romer, P. M. (1986), "Increasing Returns and Long-run Growth," *Journal of Political Economy*, 94, 1002–1037.

Schmidt, P. and R. C. Sickles (1984), "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2, 367–374.

Schwarz, G. (1978), "Estimating the Dimension of A Model," *Annals of Statistics*, 6, 461–464.

Sickles, R. C. (2005), "Panel Estimators and the Identification of Firm-Specific Efficiency Levels in Parametric, Semiparametric and. Nonparametric Settings," *Journal of Econometrics*, 126, 305–324.

Sickles, R. C. and B. Cigerli (2009), "Krugman and Young Revisited: A Survey of the Sources of Productivity Growth in a World with Less Constraints," *Seoul Journal of Economics*, 22, 29–54.

Smolny, W. (2000), "Sources of Productivity Growth: An Empirical Analysis with German Sectoral Data," *Applied Economics*, 32, 305–314.

Solow, R. M. (1957), "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39, 312–320.

Stiroh, K. (2001), "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?," *Federal Reserve Bank of New York Staff Report No. 115*.

Tsionas, E. G. (2006), "Inference in Dynamic Stochastic Frontier Models," *Journal of Applied Econometrics*, 21, 669–676.

Tullock, G. (1980), "Efficient Rent Seeking," in J. M. Buchanan, R. D. Tollison, and G. Tullock, eds., *Toward a Theory of the Rent-Seeking Society*, Texas A&M University Press, College Station, 97-112.

World Bank (2004), *2004 World Bank Indicators*, International Bank for Reconstruction and Development/The World Bank, Washington, DC.

Young, A. (1992), "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore," in *NBER Macroeconomics Annual 1992, Volume 7*, MIT Press, Cambridge, 13–64.

Young, A. (1995), "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *The Quarterly Journal of Economics*, 110, 641–680.

Zarnowitz, V. and L. A. Lambros (1992), "Consensus and Uncertainty in Economic Prediction," in *Business Cycles: Theory, History, Indicators, and Forecasting,* National Bureau of Economic Research, Inc., Cambridge, MA.

10   Appendix A

10.1   Monte Carlo Results

In this Appendix we compare the finite sample performance of weighted estimators and individual estimators through two Monte Carlo experiments. In the first study we generate simulated data following Sickles (2005) . In the second study, we construct our samples based on a data generating process that replicates the World Productivity Database from United Nation Industrial Development Organization used in our analyses.

In the first Monte Carlo experiment the base-line productivity model is:

$$Y_{it} = X'_{it}\beta - u_{it} + \varepsilon_{it}, \text{where } \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \tag{21}$$

with $\beta = (0.5, 0.5)$, $\sigma_X^2 = 1$ and $\sigma_\varepsilon^2 = 1$. In each simulated sample, the regressors are generated according to:

$$X_{it} = R X_{i,t-1} + \eta_{it}, \text{ where } \eta_{it} \sim N(0, \sigma_X^2 I_2), \ R = \begin{pmatrix} 0.4 & 0.05 \\ 0.05 & 0.4 \end{pmatrix}. \tag{22}$$

We initialize the simulation by choosing $X_{i1} \sim N(0, \sigma_X^2 (I_2 - R^2)^{-1})$, and then start iterations from $t \geq 2$. The values of regressors then are shifted around three different means $\mu_1 = (5, 5)'$, $\mu_2 = (7.5, 7.5)'$, $\mu_3 = (10, 10)'$ to obtain 3 balanced groups of firms. We $M = 1000$ samples with $n = 30, t = 30$; $n = 51, t = 21$; and $n = 21, n = 51$, according to two different DGP scenarios. The first scenario is considered as the "no problem" case: The random errors are i.i.d., the efficiency components are generated independently from a lognormal distribution and are temporally invariant, and there is no correlation between the effects and the regressors. In this scenario, samples of different groups cannot distinguish among themselves. Estimators such as CSSG and BC which does not assume correlations between the effects and the regressor should appear to have superior performance. In the second scenario, we generated η_{it} and the efficiencies from a bivariate process with 0.5 correlation. Allowing for the possible correlation between efficiency and regressors is appealing. For example, Sickles (2005) pointed out that technical efficiency in the airline industry may be due to the regressors that determine the provision of airline service such as sluggish adjustment in quasi-fixed factor such as labor in European national airlines before liberalization efforts in the late 1990s. We face inconsistent estimation if we use estimators that do address this source of endogeneity.

We report results on bias (we report the average of absolute value of bias from the true values) and mean square error (MSE). MSE is computed as:

$$MSE = \sum_{j=1}^{2} \frac{1}{M} \sum_{m=1}^{M} (\widehat{\beta}_j^m - 0.5)^2 \tag{23}$$

where $\widehat{\beta}_j^m$ are based on 6 individual methods (CSSG, EIV, BC, PSS1, PSS2W, PSS2G) and 5 different combining methods based on weights determined by: simple arithmetic average, residual sum-of-squares (RSS), R-squared, Akaike Information Criteria (AIC) and Bayes Information Criteria (BIC). We refer to this summary statistic as MSE1.

In addition to MSE 1 we can take into account correlations between different estimators in constructing the variance of the model averaged estimator. The details of the method to calculate variances for combined estimators are discussed above. We report the MSE's calculated in this way as MSE2

In Table A.1 for the no problem case, we can see that in all three simulated samples, the combined methods have smaller finite-sample biases compared to CSSG, EIV and BC. MSE1s and MSE2 calculated using the combining methods are close to each other and all of them are smaller than MSE1s and MSE2s obtained from individual methods. Table A.2 is for the scenario that regressors are correlated to the effects. Except for PSS1, which is modeled to deal with this situation, in the most cases the combined estimators have either the same or smaller magnitudes of bias than individual estimators. On the comparison of MSE1 and MSE2, combined estimators are clearly the best performers, except in one case where MSE2 of BIC is slightly larger for PSS2W and PSS2G.

To test the robustness of models in real world problems, it is reasonable practice to simulate data based on collected real data. We will construct observations of the second example following this principle.

In the second study, we compare the performance of the averaged estimators with individual frontier model estimators based on the simulated samples from the World Productivity Database of UNIDO. these results are provided in Table A.3. In the second study we use data for the developed countries (for discussions about accuracy of data, see Hulten and Isaksson (2007)). We choose K06 and EMP as capital and labor input. We generate two sets of samples utilizing a Cobb-Douglas constant returns to scale production function with $\beta = (0.5, 0.5)$ and $(0.3, 0.7)$. The different weights on β can reflect individual researcher's opinions on the contribution weights of the input factors. $(0.3, 07)$ is most commonly used selection in the literature to simulate GDP data from a Cobb-Douglas type production function. Notice that all the estimators are symmetric with respect to β except EIV. As shown in the table, in the cases of $\beta = (0.5, 0.5)$ and $\beta = (0.3, 0.7)$, all five averaging methods have smaller MSE1s and MSE2s than the individual methods, except, interestingly, for the BC estimator.

Table A.1.1			
(n = 20, t = 50)			
	Bias	MSE1	MSE2
CSSG	0.00696	0.00725	0.00726
EIV	0.00527	0.00867	0.00868
BC	0.00668	0.00565	0.00565
PSS1	0.00227	0.00700	0.00701
PSS2W	0.00042	0.00610	0.00610
PSS2G	0.00043	0.00610	0.00611
Average	0.00103	0.00308	0.00471
RSS	0.00095	0.00305	0.00478
R2	0.00123	0.00304	0.00472
AIC	0.00106	0.00312	0.00480
BIC	0.00105	0.00308	0.00472

Table A.1.2			
(n = 30, t = 30)			
	Bias	MSE1	MSE2
CSSG	0.00952	0.00883	0.00884
EIV	0.00868	0.01091	0.01092
BC	0.00420	0.00697	0.00697
PSS1	0.00146	0.00811	0.00812
PSS2W	0.00265	0.00765	0.00766
PSS2G	0.00261	0.00767	0.00768
Average	0.00299	0.00354	0.00507
RSS	0.00285	0.00346	0.00517
R2	0.00325	0.00347	0.00509
AIC	0.00306	0.00367	0.00516
BIC	0.00305	0.00355	0.00509

Table A.1.3			
(n = 50, t = 20)			
	Bias	MSE1	MSE2
CSSG	0.00775	0.00726	0.00727
EIV	0.00726	0.00989	0.00990
BC	0.00260	0.00571	0.00572
PSS1	0.00510	0.00729	0.00730
PSS2W	0.00277	0.00545	0.00546
PSS2G	0.00275	0.00546	0.00547
Average	0.00207	0.00307	0.00441
RSS	0.00203	0.00301	0.00449
R2	0.00231	0.00303	0.00442
AIC	0.00209	0.00314	0.00449
BIC	0.00216	0.00309	0.00443

Table A.2.1			
(n = 20, t = 50)			
	Bias	MSE1	MSE2
CSSG	0.00338	0.00152	0.00152
EIV	0.00280	0.00167	0.00167
BC	0.00410	0.00162	0.00163
PSS1	0.00166	0.00231	0.00231
PSS2W	0.00283	0.00145	0.00145
PSS2G	0.00283	0.00145	0.00145
Average	0.00293	0.00076	0.00109
RSS	0.00305	0.00076	0.00110
R2	0.00298	0.00077	0.00109
AIC	0.00279	0.00078	0.00111
BIC	0.00322	0.00086	0.00126

Table A.2.2			
(n = 30, t = 30)			
	Bias	MSE1	MSE2
CSSG	0.00497	0.00176	0.00176
EIV	0.00401	0.00199	0.00199
BC	0.00664	0.00172	0.00172
PSS1	0.00109	0.00283	0.00283
PSS2W	0.00395	0.00187	0.00188
PSS2G	0.00395	0.00187	0.00188
Average	0.00394	0.00092	0.00131
RSS	0.00394	0.00093	0.00132
R2	0.00397	0.00093	0.00132
AIC	0.00392	0.00093	0.00134
BIC	0.00509	0.00106	0.00163

Table 2.2.3			
(n = 50, t = 20)			
	Bias	MSE1	MSE2
CSSG	0.00810	0.00160	0.00161
EIV	0.00545	0.00183	0.00183
BC	0.01038	0.00165	0.00165
PSS1	0.00033	0.00261	0.00261
PSS2W	0.00630	0.00147	0.00147
PSS2G	0.00631	0.00147	0.00147
Average	0.00603	0.00083	0.00120
RSS	0.00609	0.00084	0.00112
R2	0.00606	0.00083	0.00112
AIC	0.00597	0.00083	0.00113
BIC	0.00855	0.00114	0.00181

	Table A.3					
(LnK, LnL)	(0.3, 0.7)			(0.5, 0.5)		
	Bias	MSE1	MSE2	Bias	MSE1	MSE2
CSSG	0.00169	0.01400	0.01401	0.00169	0.01400	0.01400
EIV	0.02369	0.06781	0.06925	0.00754	0.06781	0.06781
BC	0.00477	0.00067	0.00067	0.00479	0.00067	0.00067
PSS1	0.00226	0.02572	0.02575	0.00226	0.02572	0.02575
PSS2W	0.00068	0.01333	0.01334	0.00068	0.01333	0.01334
PSS2G	0.00100	0.01327	0.01327	0.00100	0.01326	0.01327
Average	0.00113	0.00704	0.00944	0.00114	0.00704	0.00944
RSS	0.00098	0.00539	0.00944	0.00098	0.00539	0.00836
R2	0.00114	0.00695	0.00836	0.00114	0.00695	0.00938
AIC	0.00108	0.00625	0.00938	0.00108	0.00419	0.00891
BIC	0.00118	0.00707	0.00891	0.00118	0.00707	0.00947

Table 1: Combined Estimates Result Presentation

		Arithmetic	RSS	R-Square	AIC	BIC
Low	Estimate	0.0047	0.0050	0.0046	0.0047	0.0033
	Variance	5.32E-05	1.63E-05	5.02E-05	3.69E-05	4.23E-05
	Bound	1.74E-04	8.84E-05	1.67E-04	1.37E-04	1.38E-04
Low-Mid	Estimate	0.0032	0.0012	0.0030	0.0020	0.0023
	Variance	5.78E-05	5.81E-06	5.48E-05	2.77E-05	4.53E-05
	Bound	2.19E-04	8.37E-05	2.12E-04	1.48E-04	1.69E-04
Upper-Mid	Estimate	0.0079	0.0066	0.0079	0.0071	0.0075
	Variance	1.79E-05	5.32E-06	1.77E-05	1.18E-05	1.56E-05
	Bound	1.52E-04	1.10E-05	1.51E-04	1.34E-04	1.32E-04
High	Estimate	0.0051	0.0040	0.0049	0.0037	0.0035
	Variance	6.86E-06	4.32E-06	6.15E-06	2.71E-06	4.41E-06
	Bound	7.47E-05	6.25E-05	7.33E-05	6.66E-05	3.84E-05
Old Tigers	Estimate	0.0393	0.0367	0.0391	0.0392	0.0410
	Variance	1.14E-04	3.15E-04	1.08E-04	1.16E-04	2.51E-04
	Bound	1.30E-03	1.33E-03	1.29E-03	1.32E-03	1.45E-03
New Tigers	Estimate	0.0286	0.0331	0.0287	0.0309	0.0283
	Variance	5.29E-05	2.43E-04	3.35E-05	6.87E-05	6.09E-05
	Bound	8.32E-04	1.09E-03	8.41E-04	9.71E-04	8.15E-04